

Controlling for Endogeneity between Bus Headway and Bus Ridership: A Case Study of the Orlando Region

Moshiur Rahman

Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 321-276-7580, Fax: 407-823-3315
Email: moshiur@knights.ucf.edu
ORCID number: 0000-0002-3210-6542

Shamsunnahar Yasmin*

Research Fellow - Road Safety Engineering
Centre for Accident Research and Road Safety – Queensland
Queensland University of Technology
Tel: 61 7 3138 4677; Fax: 61 7 3138 7532
Email: shams.yasmin@qut.edu.au

Naveen Eluru

Associate Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: naveen.eluru@ucf.edu
ORCID number: 0000-0003-1221-4113

ABSTRACT

In this study, we develop an advanced econometric model that considers the potential endogeneity of stop level headway in modeling bus ridership. We recognize that bus stops with higher potential demand are also like to have higher frequency of buses (or lower headway between buses). We consider headway endogeneity by proposing a simultaneous equation system that considers headway and ridership in a joint framework. The proposed model is developed employing stop level ridership data from the Orlando region for 11 quadrimesters (four-month time periods). The presence of multiple data points for each stop allows us to develop panel models for headway, boarding, and alighting. The headway variable is modeled using a panel ordered logit model while the ridership variables are modeled using a panel grouped ordered logit models. The model estimation results justify the consideration of headway endogeneity in bus ridership analysis. To illustrate the value of the proposed model, a validation exercise and a policy analysis exercise are conducted.

Keywords: *Endogeneity; Headway; Joint Framework; Panel Ordered Logit; Panel Grouped Order Logit*

INTRODUCTION

In urban regions, public transportation systems ought to provide an equitable, safe and accessible transportation mode for residents. According to 2016 American Community Survey data, transit mode only accounts for about 5% of the commute trips in the United States (Tomer 2016). Existing public transportation systems are either facing reduction in ridership and/or facing challenges with regards to providing equitable services to residents. In fact, in recent years, several urban transit systems have experienced declines in ridership (Gomez-Ibanez 1996, Garrett and Taylor 1999, Bliss 2017, Schmitt 2017, Lewyn 2018, Siddiqui 2018). Ideally, in the presence of a well-designed public transit system, urban residents irrespective of their ethnicity, household income, and vehicle ownership should have similar access to activity participation or employment opportunities. Several researchers have found evidence to the contrary while examining the influence of transportation on employment opportunities (Shen 2001, Wenglenski and Orfeuill 2004, Kawabata and Shen 2006, 2007, Grengs 2010, Boarnet *et al.* 2017). These studies identified that access to employment by transit is substantially lower than access to employment by car mode. However, several public transit riders own no cars and are reliant on public transportation to arrive at work. Thus, there is a need to examine public transportation system design and operation to enhance transit adoption and equity for urban residents.

Policy makers and urban agencies across different parts of North America, are considering investments in various public transportation alternatives including bus, light rail, commuter rail, and metro (see Jaffe 2014, TP 2016 for public transportation projects under construction or consideration). A critical component for evaluating the success of these investments is the development of appropriate statistical tools to examine the impact. Our proposed research contributes to public transit literature by developing econometric models that consider the potential endogeneity of stop level headway in modeling ridership. To elaborate, earlier research in public transportation has identified headway (alternatively bus frequency) as one of the primary determinants affecting ridership. The studies conclude that stops with higher headway (lower frequency) between buses are likely to have lower ridership. While this is a perfectly acceptable conclusion, most (if not all) studies in public transit literature ignore that the stop level headway was determined (by choice) in response to expected ridership i.e. stops with lower headway were expected to have higher ridership numbers. In traditional ridership studies, this potential endogeneity is often neglected, and headway is considered as an independent variable. The approach violates the requirement that the unobserved factors affecting the dependent variable do not have impact on independent variables. If this is the case, the estimated impact of headway on ridership would be biased (potentially over-estimated). More importantly, the estimated impact of all other variables (such as land use factors, bus infrastructure) will also be biased (possibly under-estimated). Traditional ridership models also consider transit ridership at a single time point for analysis using cross-sectional datasets. Ideally, it would be beneficial to consider data from multiple time points. The consideration of data from multiple time points is of particular value in accommodating for the impact of headway associated endogeneity.

In this study, we address these challenges by proposing a simultaneous equation system that considers headway and ridership (characterized as boarding and alighting) in a joint framework while accounting for the influence of common unobserved factors affecting headway and ridership. The proposed model is developed employing ridership data from Orlando region for the Lynx bus transit system. The ridership data includes stop level quadrimester (four-month time period) average weekday boarding and alighting information. The average ridership information are available for 11 quadrimesters from May 2013 to December 2016. The presence of multiple

data points for each stop allows us to develop panel models for headway, boarding and alighting. In the joint modeling approach, the headway variable is modeled using a panel ordered logit model while the ridership variables are modeled using panel grouped ordered logit models. In addition to unobserved effects in the form of panel random effects, several exogenous variables including stop level attributes (such as number of bus stop), transportation infrastructure variables (such as secondary highway length, rail road length and local road length, sidewalk length), transit infrastructure variables (such as bus route length, presence of shelter and distance of bus stop from central business district (CBD)), land use and built environment attributes (such as land use mix, residential area, recreational area, institutional area, office area, etc.) and demographic and socioeconomic variables in the vicinity of the bus stop (income, vehicle ownership, age and gender distribution) are considered in the model estimation. Finally, to illustrate the model applicability, we generate changes to ridership based on changes in multiple independent variables.

The remainder of the paper is organized as follows. A brief overview of earlier research is described in the literature review section. The methodology section outlines the econometric frameworks considered. The data section presents data source and data preparation for analysis. The model estimation results section presents the estimation results and validation. The policy analysis results are discussed in the next section. Finally, the conclusion section provides a summary of the findings and concludes our paper.

LITERATURE REVIEW

Background

Transit ridership literature can be categorized into two groups. The first group of studies focuses on the factors that affect transit adoption at a disaggregate level by exploring individual perceptions and behavioral responses (Fan *et al.* 1993, Handy 1996, Wardman and Whelan 1999, Balcombe *et al.* 2004, Evans 2004, McCollom and Pratt 2004a, McCollom and Pratt 2004b, Handy *et al.* 2005, Debrezion *et al.* 2007, 2009, Van Acker *et al.* 2010, Chakour and Eluru 2014). The second group of studies examines the impact of various factors on system level (or route level) ridership measures (Seskin *et al.* 1996, FitzRoy and Smith 1998, Kain and Liu 1999, Babalik-Sutcliffe 2002, Johnson 2003, Mackett and Sutcliffe 2003, Ma *et al.* 2018). The current proposed research effort falls into the second group of studies. A detailed review of all these studies is beyond the scope of the paper. The reader is referred to a recent study by Rahman *et al.* (2019) that provides a detailed summary of literature across these two groups. In this section, we focus on literature particularly relevant to our research effort. We begin with an overview of studies in transportation that attempt to accommodate for endogeneity. Subsequently, we examine studies that consider endogeneity within transit literature.

Addressing Endogeneity in Transportation

The travel behavior field has extensively examined the influence of endogeneity across various decision processes. Specifically, these studies have explored the potential impact of residential location choice – labelled as residential self-selection – on various travel behavior choices (Bhat and Guo 2007, Mokhtarian and Cao 2008, Bhat and Eluru 2009, Cao *et al.* 2009, Pinjari *et al.* 2009, Walker *et al.* 2011, Aditjandra *et al.* 2012, Vij and Walker 2014, Ding *et al.* 2017, Ettema and Nieuwenhuis 2017). More recently, several researchers developed/analyzed approaches to incorporate endogeneity associated with omitted variables such as attitudinal attributes (Paulsen *et al.* 2014, Fernández-Antolín *et al.* 2016, Guevara and Polanco 2016, Guevara 2015, Beck *et al.*

2017, Mariel *et al.* 2018). Furthermore, several studies have incorporated the impact of individual attitudes and social influence characteristics on travel behavior within hybrid choice models (Chorus and Kroesen 2014, Kamargianni *et al.* 2014, Kim *et al.* 2014, Kim *et al.* 2017).

There are examples from other fields including seat belt choice in driver injury severity models (Eluru and Bhat 2007, Abay *et al.* 2013); emergency medical response time affecting fatality timeline (Yasmin *et al.* 2015) and bicycle sharing system station capacity decision influencing bicycle sharing demand (Faghih-Imani and Eluru 2016). The endogenous variables and the choice variables could be examined as continuous or discrete indicators. Based on the nature of the variables involved, several approaches such as proxy variables (Wardman and Whelan 2011, Tirachani *et al.* 2013), control function method (Guevara and Ben-Akiva 2012), multiple indicator solution (Guevara and Polanco 2016), instrument variables regression, two-stage residual inclusion (2SRI) approach and Roy's (Roy 1951) endogenous system or the treatment effects model (see Maddala 1983, Heckman and Vytlacil 2005), and joint econometric modeling approaches (Eluru and Bhat 2007) are employed.

Research in Transit Field Accommodating Endogeneity

Given the prevalence of modeling approaches for addressing endogeneity bias in transportation field, it is not surprising that multiple studies have either alluded to the presence of endogeneity or specifically employed approaches to control for it in the context of public transit analysis. Earlier research in transit ridership analysis have discussed potential endogeneity of transit ridership and transit price, service and automobile ownership dimensions (Creutzig 2014). Holmgren (2007) conducted a meta-analysis of elasticity estimates of bus demand in transit literature and recommended that service variable (headway) should be treated as endogenous while other variables such as car ownership, fuel price and ticket price should be considered as exogenous variables. The studies that considered endogeneity have controlled for different dimensions governed by the authors' judgement. Voith (1991) developed community transit demand models while accommodating for the interaction between transit fare prices and service decisions on ridership. The authors estimate a dynamic fixed effects panel model with Instrumental Variables (IV) using data from Southeastern Pennsylvania Transportation Authority (SEPTA). Voith (1997) extended the model developed in Voith (1991) with a larger data sample with IV approach developing separate equations for price and service.

Fitzroy and Smith (1999) developed a framework to examine the impact of season tickets on transit ridership across four Swiss cities. To account for the potential impact of investments on road and transit infrastructure on overall ridership, the authors employed an IV approach. Further, the authors controlled for potential contemporaneous unobserved correlation by developing seemingly unrelated regression approach. Deka (2002) examined the potential endogeneity of automobile ownership and transit availability in the Los Angeles region. Specifically, the author estimated a model for transit availability and employed its predicted value as an independent variable in modeling automobile ownership. Novak and Savage (2013) studied the cross-elasticity between fuel price and transit usage for the Chicago region for various rail and bus services. The authors indicate that adopting a two stage least squares approach leads to counter-intuitive results in their data analysis. The reader would note that a majority of these studies develop models at a system level i.e. employ aggregate measures of ridership.

Current Study in Context

The literature review highlights how well recognised the issue of endogeneity is within the transit filed. However, the literature is not without limitations. First, while several studies have explicitly considered/controlled for endogeneity, the study frameworks focus on aggregate transit ridership metrics such as monthly boardings at the system level. There is no study that has examined the endogeneity issue at a more disaggregate level such as bus route or stop level. The aggregate level models are adequate for planning at a system level. However, for any analysis of changes to the existing service for various bus routes, more detailed analysis at the bus route or stop level is warranted. Second, earlier analysis was explored using cross-sectional or panel data with very small data samples. This is expected because the analysis was conducted at a system level yielding smaller data samples. Third, while several studies developed IV and/or 2SRI approaches there is no effort in the discrete choice realm controlling for endogeneity. The current research effort addresses these limitations by undertaking a disaggregate stop level ridership analysis (for boarding and alighting) while controlling for endogeneity associated with stop-level headway. For the Orlando region, while headway is a continuous value in minutes, due to the nature of the service in the region, it is more accurate to consider headway as a discrete variable. In our study, we have considered three categories for headway model: (i) Category 1 (0-15 minutes), (ii) Category 2 (15-30 minutes) and (iii) Category 3 (>30 minutes). Hence, we have considered headway as an ordered discrete variable. Further, to model ridership, building on our earlier work (Rahman *et al.* 2019), we categorize the boardings and alightings as grouped ordered variables¹. Thus, the overall econometric methodology employed results in a panel multivariate ordered system with three separate equations (for headway, boarding and alighting). The proposed model system is estimated using data for 11 quadrimesters (four-month periods) from May 2013 to December 2016. The proposed joint panel modeling approach is the first of its kind for transit ridership analysis to the best of the authors' knowledge.

METHODOLOGY

The focus of this study is to examine stop-level boarding, alighting and headway simultaneously. Let q ($q = 1, 2, \dots, Q$) be an index to represent bus stops, t ($t = 1, 2, 3, \dots, T$) represent the different time periods, m ($m = 1, 2, \dots, M=3$) be an index to represent headway categories and j ($j = 1, 2, 3, \dots, J = 13$) be an index to represent the categories of boardings or alightings. For headway component, we consider three categories: category 1 = 0-15 minutes; category 2 = 15-30 minutes and category 3 = > 30 minutes. The thirteen categories for ridership analysis considered are: Bin 1 = ≤ 5 ; Bin 2 = 5-10; Bin 3 = 10-20, Bin 4 = 20-30, Bin 5 = 30-40, Bin 6 = 40-50, Bin 7 = 50-60, Bin 8 = 60-70, Bin 9 = 70-80, Bin 10 = 80-90, Bin 11 = 90-100, Bin 12 = 100-120 and Bin 13 = >120. Then, the equation system for modeling headway, boarding and alighting jointly can be written as:

$$H_{qt}^* = (\nu' + \sigma_q')x'_{qt} + (\eta'_q)y_{qt} + \Delta_{qt}, H_{qt} = m \text{ if } \varpi_{m-1} < H_{qt}^* \leq \varpi_m \quad (1)$$

¹ The grouped response approach proposed in Rahman et al. (2019) provides a true non-linear variant of the linear regression model structure (with the same number of model parameters as linear regression). In this model system, while continuous data is binned into various categories, the binning is simply to facilitate model estimation. Unlike traditional ordered discrete models, the grouped response model does not estimate thresholds in the ridership model. The thresholds of the bins selected are directly used in the model estimation; thus, observed levels of the dependent variable are tied to the propensity. Post model estimation, the model developed can be used to estimate the probability of any ridership level (see Rahman et al. 2019 for more details).

$$\begin{aligned}
B_{qt}^* &= (\alpha' + \gamma'_q)x''_{qt} + (\theta' + \mu'_q)h_{qt} \pm (\eta'_q)y_{qt} + \varepsilon_{qt}; \\
B_{qt} &= j \text{ if } \psi_{j-1} < B_{qt}^* \leq \psi_j
\end{aligned} \tag{2}$$

$$\begin{aligned}
A_{qt}^* &= (\beta' + \delta'_q)x''_{qt} + (\theta'' + \mu''_q)h_{qt} \pm (\eta'_q)y_{qt} + \xi_{qt}; \\
A_{qt} &= j \text{ if } \psi_{j-1} < A_{qt}^* \leq \psi_j
\end{aligned} \tag{3}$$

In equation 1, H_{qt}^* is the latent (continuous) propensity for headway at stop q for the t^{th} time period. This latent propensity H_{qt}^* is mapped to the actual headway category m by the ϖ thresholds, in the usual ordered-response modeling framework. x'_{qt} is a matrix of attributes that influences stop level headway, ν is the vector of mean coefficients and σ_q is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of x'_{qt} .

In equations 2 and 3, B_{qt}^* (A_{qt}^*) is the latent propensity for stop level boardings (alightings) of stop q for the t^{th} time period. This latent propensity B_{qt}^* (A_{qt}^*) is mapped to the actual grouped ridership category j by the ψ thresholds, in the usual ordered-response modeling framework. In our case, we consider $J = 13$ and the ψ values are fixed as follows: $-\infty, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120$, and $+\infty$. x''_{qt} is a matrix of attributes that influences stop level boarding and alighting. α (β) is the corresponding vector of mean coefficients and γ_q (δ_q) is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of x''_{qt} for boardings (alightings), h_{qt} represents the headway variables generated from H_{qt} for consideration in boarding and alighting. θ' (θ'') represents the corresponding vector of mean coefficients and μ'_q (μ''_q) is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element h_{qt} for boardings (alightings). ε_{qt} (ξ_{qt}) is an idiosyncratic random error term assumed to be independently logistic distributed across choice stops and choice occasions for boardings (alightings) with variance λ_B^2 (λ_A^2). The variance vectors for boarding's and alighting's are parameterized as a function of independent variables as follows: $\lambda_B = \exp(\pi' l_{qt})$ and: $\lambda_A = \exp(\pi'' l_{qt})$ where l_{qt} represent independent variables that influence the variance, while π' and π'' are corresponding coefficient vectors. The parameterization allows for the variance to be different across the bus stops accommodating for heteroscedasticity. η_q present in all three equations represents the vector of coefficients that accommodates for the impact of stop level common unobserved factors that jointly influence boardings, alightings and headway. The ' \pm ' sign indicates that the potential impact could be either positive or negative. A positive sign implies that unobserved factors that increase the headway for a given reason will also increase the propensity for boarding/alighting, while a negative sign suggests that unobserved individual factors that increase the propensity for headway will decrease the propensity for boarding/alighting. In our empirical context, we expect the relationship to be negative.

Further, to accommodate for ridership category specific effects z_{qjt} is a vector of attributes specific to stop q and ridership category alternative j , while ρ_j and τ_j is the vector of corresponding ridership category-specific coefficients for boarding and alighting components, respectively. To complete the model structure of Equations (1), (2) and (3), it is necessary to define the structure for the unobserved vectors $\gamma_q, \delta_q, \mu_q$ (combined vector of μ'_q and μ''_q) and η_q . In this paper,

we assume that these vectors are independent realizations from normal distributions as follows: $\gamma_{qn} \sim N(0, \kappa_n^2)$, $\delta_{qn} \sim N(0, \nu_n^2)$, $\sigma_{qn} \sim N(0, \varsigma_n^2)$, $\mu_{qn} \sim N(0, o_n^2)$ and $\eta_{qn} \sim N(0, \varrho_n^2)$.

With these assumptions, the probability expressions for the ridership category may be derived. Conditional on γ_{qn} , δ_{qn} , σ_{qn} , μ_{qn} and η_{qn} , the probability for stop q to have headway, boarding, and alighting in the t^{th} time period is respectively given by:

$$P(H_{mt})|\sigma, \eta = \Lambda \left[\varpi_m - \left((\nu' + \sigma'_q) x'_{qt} + (\eta'_q) y_{qt} \right) \right] - \Lambda \left[\varpi_{m-1} - \left((\nu' + \sigma'_q) x'_{qt} + (\eta'_q) y_{qt} \right) \right] \quad (4)$$

$$P(B_{jt})|\gamma, \eta = \Lambda \left[\frac{\psi_j - \left((\alpha' + \gamma'_q) x''_{qt} + (\rho'_j) z_{qjt} + (\theta' + \mu'_q) h_{qt} \pm (\eta'_q) y_{qt} \right)}{\lambda_B} \right] - \Lambda \left[\frac{\psi_{j-1} - \left((\alpha' + \gamma'_q) x''_{qt} + (\rho'_j) z_{qjt} + (\theta' + \mu'_q) h_{qt} \pm (\eta'_q) y_{qt} \right)}{\lambda_B} \right] \quad (5)$$

$$P(A_{jt})|\delta, \eta = \Lambda \left[\frac{\psi_j - \left((\beta' + \delta'_q) x''_{qt} + (\tau'_j) z_{qjt} + (\theta'' + \mu''_q) h_{qt} \pm (\eta'_q) y_{qt} \right)}{\lambda_B} \right] - \Lambda \left[\frac{\psi_{j-1} - \left((\beta' + \delta'_q) x''_{qt} + (\tau'_j) z_{qjt} + (\theta'' + \mu''_q) h_{qt} \pm (\eta'_q) y_{qt} \right)}{\lambda_B} \right] \quad (6)$$

where $\Lambda(\cdot)$ is the cumulative standard logistic distribution.

Let Ω represent a vector that includes all the standard error parameters to be estimated. Given these assumptions the joint likelihood for stop level boarding and alighting is provided as follows:

$$L_q|\Omega = \prod_{t=1}^T \left[\prod_{m=1}^M [P(H_{mt})|\sigma, \eta]^{d_{hmt}} * \left\{ \prod_{j=1}^J [(P(B_{jt})|\gamma, \eta)]^{d_{bjt}} [(P(A_{jt})|\delta, \eta)]^{d_{ajt}} \right\} \right] \quad (7)$$

where d_{hmt} is a dummy variable taking a value of 1 if stop q has headway within the m^{th} category for the t^{th} time period and 0 otherwise; d_{bjt} and d_{ajt} are dummy variables taking a value of 1 if stop q has ridership within the j^{th} category for the t^{th} time period and 0 otherwise. Finally, the unconditional likelihood function may be computed for stop q as:

$$L_q = \int_{\Omega} (L_q|\Omega) f(\Omega) d\Omega \quad (8)$$

The log-likelihood function is given by

$$\text{Ln}(L) = \sum_{q=1}^Q \ln L_q \quad (9)$$

The likelihood function in Equation (8) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in Ω . We apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function (See Bhat 2003, Yasmin and Eluru 2013 for more details). The likelihood functions are programmed in GAUSS (Aptech 2016).

EMPIRICAL ANALYSIS

The current research effort is conducted using data from the Greater Orlando region. Orlando is a typical American city in the south with the following transportation mode share: automobile (85.7%), public transit (1.0%), walk (9.2%) and bike (1.2%). The main public transit system serving the Orlando metropolitan region is the Lynx transit system. Lynx system serves the population of about 1.8 million in Orange, Seminole, Osceola and Polk County covering 2,500 square miles. The system has 77 daily routes offering about 105,682 rides on weekdays. The number of bus stops considered for analysis include 3,444 stops. Of these, 2,800 stops data are used for model estimation while 644 stops data are set aside for validation from each quarter. The average weekday ridership data for each time period were obtained from Lynx transit authority. For our analysis, headway, average weekday boarding and alighting ridership data was considered from May 2013 to December 2016 for 11 quadrimesters. The final sample consists of 37,884 records (3,444 stops \times 11 quarters).

Dependent Variables

For our analysis, headway variable and ridership variables – boarding and alighting – are considered as dependent variables. For headway variable, the percentage across categories is 9.1%, 37.7% and 53.2% for category 1, category 2 and category 3, respectively. The average daily stop level boarding (alighting) is around 18.84 (18.70) with a minimum of 0 (0) and maximum of 6,135 (5,943). A summary of the system-level ridership (boarding and alighting) is provided in Table 1. The reader would note that the standard deviation of ridership reported is large as the ridership varies significantly across different bus stops.

Table 1 Summary Statistics of Lynx Bus Ridership (August 2013 to December 2016)

Time-period	Quarter Name	Boarding		Alighting	
		Mean	Standard Deviation	Mean	Standard Deviation
1	August-13	19.91	140.54	19.63	132.67
2	December-13	19.17	135.70	19.04	129.16
3	April-14	19.03	142.17	18.88	137.42
4	August-14	19.66	144.18	19.50	136.68
5	December-14	18.51	132.80	18.45	128.70
6	April-15	18.81	138.54	18.89	133.20
7	August-15	18.79	138.63	18.77	132.55
8	December-15	18.55	131.09	18.43	129.42

9	April-16	17.84	127.10	17.83	126.67
10	August-16	18.64	131.77	18.50	130.15
11	December-16	18.29	129.38	17.84	124.80

Independent Variables

Several exogenous variables are generated to augment the ridership information. The information is sourced from Lynx GIS shapefiles, 2010 US census data, American Community Survey, Florida Geographic Data Library, and Florida Department of Transportation (FDOT) databases. The exogenous attributes considered in our study can be divided into four broad categories: (1) Stop level attributes (such as headway), (2) Transportation and transit infrastructure variables (secondary highway length, rail road length and local road length, sidewalk length, Lynx bus route length, presence of shelter and distance of bus stop from central business district (CBD)), (3) Built environment and land use attributes (such as institutional area, residential area, recreation area, office area) (4) Demographic and socioeconomic variables in the vicinity of the stop (such as income, vehicle ownership, and age and gender distribution) for each time period. For generating exogenous variable values, we have considered several buffer distances (800m, 600m, and 400m) around each bus stop. The descriptive statistics of exogenous variables are presented in Table 2.

MODEL ESTIMATION RESULTS

Model Specification and Overall Measures of Fit

The empirical analysis involves estimation of different models: 1) Independent ridership-headway (IRH) model that does not accommodate for headway endogeneity and 2) Joint ridership-headway (JRH) model that explicitly accommodates for headway endogeneity. Prior to discussing the estimation results, we compare the performance of these models in this section. We employ the Bayesian Information Criterion (BIC) to determine the best model between independent and joint model. The BIC for a given empirical model is equal to:

$$BIC = -2LL + K \ln(Q) \quad (10)$$

where LL is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lower BIC is the preferred model. The log-likelihood values at convergence for the models estimated are as follows: (1) IRH model (with 54 parameters) is -110,705.364 (2) JRH model (with 56 parameters) is -104,965.476. The BIC values for the final specifications of IRH and JRH are 221,968.833 and 210,509.727, respectively. The comparison exercise clearly highlights the superiority of the Joint model in terms of data fit compared to independent model.

Variable Effects

In presenting the effects of the exogenous variables, we will restrict ourselves to the discussion of the joint model². The final specification of the model was based on removing the statistically insignificant variables in a systematic process based on statistical confidence (95% confidence level). The specification process was also guided by prior research and parsimony considerations. In estimating the models, several functional forms and variable specifications are explored. The

² Estimation results for independent model is presented in Appendix A.

functional form that provided the best result was used for the final model specifications. In determining the appropriate buffer sizes, each variable for a buffer size was systematically introduced (starting from 800m to 400m buffer size) and the buffer variable that offered the best fit was considered in the final model specification.

Table 3 presents the estimation results of the joint model. Specifically, columns 2 and 3 provide the variable impacts of the headway component while columns 4 through 7 present the results of boarding and alighting components. The results are organized in 4 categories: (a) independent variables, (b) alternative specific effects for ridership component, (c) variance parameters for ridership component and (d) joint common unobserved factors. The reader would note that after accounting for joint common unobserved factors across the three variables, the impact of other unobserved factors - parameter vectors $\gamma_q, \delta_q, \sigma_q, \mu_q$ - in the headway and ridership models were not statistically significant. The model results are discussed separately for headway and ridership components.

Headway Component

The positive (negative) coefficient corresponds to increased (decreased) propensity for longer headway categories.

Transportation Infrastructure Characteristics

The bus route length of 800m buffer has a negative impact on headway. The variable impact is expected. Bus stops with larger bus route length are likely to have higher frequency of bus arrivals i.e. lower headway. A negative impact of the presence of bike length in 800m vicinity of the bus stop on headway is also along expected lines. The presence of bicycle infrastructure serves as a proxy for denser neighborhoods encouraging non-automobile alternatives. The presence of increased secondary highway length in the 800m buffer decreases the headway while a corresponding increase in local road length increases headway. The roadway length variable is possibly serving as an indicator of urban locations. The results also indicate that in the presence of a railroad, headway is likely to be lower. The result warrants further investigation.

Built Environment Attributes

The built environment around a bus stop has a significant impact on bus frequency. The presence of industrial and residential areas within an 800m buffer of a bus stop is likely to increase the headway. On the other hand, in the presence of institutional, recreational and office area (800m buffer), the headway is likely to be lower. The results are intuitive. An increase in the stop distance from the CBD is likely to increase the headway (as expected).

Demographic and Socioeconomic Characteristics

In terms of demographic and socioeconomic variables, vehicle ownership variable has a significant impact. Specifically, locations with higher proportion of households with no vehicle are likely to have a lower headway value. The result is perhaps indicating the fact that households with no vehicles are captive to transit mode.

Table 2 Descriptive Statistics of Exogenous Variables

Variable Name	Variable Description	Percentage	Minimum	Maximum	Mean
Stop Level Attributes					
Dummy for headway category 1	Headway 0~15 minutes	9.094%	-	-	-
Dummy for headway category 2	Headway >15~30 minutes	37.688%	-	-	-
Dummy for headway category 3	Headway >30 minutes	53.218%	-	-	-
Number of bus stop (800m buffer)	No of bus stop within 800m buffer of a stop/10	-	0.100	9.300	1.727
Transportation Infrastructure Around the Stop					
Bus route length in a 800m buffer	Bus route length in kilometers (Bus route length in 800 m buffer/10)	-	0.000	8.710	0.878
Sidewalk length in an 400m buffer	Sidewalk length in kilometers in 400m buffer	-	0.000	7.557	0.985
Bike Lane Length (800m buffer)	Bike Lane length in km in 800m buffer	-	0.000	9.100	0.458
Secondary highway length (800m buffer)	Secondary highway length in 800 m buffer / 10	-	0.000	4.278	0.964
Railroad length in an 800m buffer	Railroad length in kilometers in 800m buffer	-	0.000	6.312	0.301
Local road length in an 800m buffer	Local road length in 800 m buffer / 10	-	0.000	6.048	2.138
Presence of shelter in bus stop	Shelter (1 = Yes and 0 = No)	22.750%	-	-	-
Built Environment Around the Bus Stop					
Industrial area (800m buffer)	Proportion of the industrial area in 800m buffer = Industrial area/Total area	-	0.000	0.738	0.054
Residential area (800m buffer)	Proportion of the Residential area in 800m buffer = Residential area/Total area	-	0.000	0.992	0.443
Institutional area (800m buffer)	Proportion of the Institutional area in 800m buffer = Institutional area/Total area	-	0.000	0.720	0.041
Recreational area (800m buffer)	Proportion of the Recreational area in 800m buffer = Recreational area/Total area	-	0.000	0.557	0.012
Office area (800m buffer)	Proportion of the office area in 800m buffer = Office area/Total area	-	0.000	0.957	0.171
Central business district (CBD) distance	(CBD distance in km from bus stop)/10	-	0.003	5.058	1.183
Demographic and Socioeconomic Variables in Census Tract					
Zero vehicle in household (HH)	Percentage of zero vehicle HH	-	0.000	0.618	0.061
High income (>100k)	Percentage of High income HH (>100k)	-	0.000	0.695	0.131
Household rent	Percentage of rented HH	50.058%	-	-	-

Table 3 Lynx Ridership Analysis Results

Variable Name	Headway		Alighting		Boarding	
	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
Independent Variables (v, α, β)						
Constant	-	-	-13.010	-16.049	-23.992	-24.242
Threshold Value 1	-3.998	-76.179	-	-	-	-
Threshold Value 2	0.302	6.016	-	-	-	-
<i>Stop Level Attributes</i>						
Headway (Base: Category 1)						
<i>Dummy for headway category 2</i>	-	-	-45.477	-99.782	-51.083	-99.457
<i>Dummy for headway category 3</i>	-	-	-70.976	-135.682	-77.723	-130.697
No of Bus stop in an 800 m buffer	-	-	-4.422	-29.238	-4.447	-26.104
Bus route Length in an 800 m buffer	-0.815	-73.068	-2.374	-15.692	-3.546	-21.583
Presence of shelter in bus stop	-	-	19.981	75.939	34.953	111.848
<i>Transportation Infrastructure</i>						
Sidewalk length in an 400 m buffer	-	-	2.694	14.902	2.184	10.767
Bike road length in an 800 m buffer	-0.204	-26.582	-	-	-	-
Secondary road length in an 800 m buffer	-0.506	-38.451	8.436	39.398	7.011	29.586
Local road length in an 800 m buffer	0.312	21.616	4.568	22.290	4.788	19.872
Railroad length in an 800 m buffer	-0.633	-53.374	-	-	-	-
<i>Built Environment and Land-use Attributes</i>						
Land use area type in an 800m buffer						
<i>Institutional area</i>	-1.879	-18.141	24.718	13.492	6.032	2.721
<i>Residential area</i>	1.770	31.501	-	-	14.806	17.950
<i>Office area</i>	-2.164	-27.960	40.455	43.718	42.40	32.023
<i>Recreational area</i>	-0.432	-1.917	-75.918	-25.522	-65.009	-19.079
<i>Industrial Area</i>	5.208	42.210	-	-	-	-
Distance from Central business district (CBD)	0.489	44.374	-3.358	-17.764	-3.796	-18.357
<i>Demographic and Socioeconomic Variables</i>						
Zero vehicle in HH	-2.206	-14.453	76.737	29.154	72.871	24.309
High income population	-0.620	-8.667	-	-	-	-
Household rent	-	-	30.784	47.983	35.120	48.800
<i>SunRail Effects</i>						
Distance Decay Function for SunRail*SunRail operation period	-	-	-5.008	-18.494	-5.167	-17.671
Alternative Specific Effects (ρ_j and τ_j)						
Constant – Alternative 1 (0-5 ridership)	-	-	37.502	125.131	42.933	123.530

Constant – Alternative 2 (>5-10 ridership)	-	-	17.910	82.816	20.340	82.613
Variance (λ_B^2 and λ_A^2)						
Constant	-	-	3.263	753.061	3.348	706.331
Joint Common Unobserved Factors (η'_q)						
Constant	1.736			157.142		
Route Length in 800m buffer	0.792			106.031		

Boarding and Alighting Components

Stop Level Attributes

In our ridership models, headway variable was considered with lowest headway category as the base case. The corresponding parameter estimates for headway categories 2 and 3 clearly indicate that increasing headway has a negative impact on ridership (for both boarding and alighting). The reader would note that incorporating headway endogeneity does not eliminate the impact of headway on ridership. It is estimated more accurately. The parameter for the number of bus stops in an 800m buffer indicates that with increasing number of stops in the buffer, ridership values decrease indicating competition across the stops (see Rahman *et al.* 2019 for similar result). The result is further reinforced by the parameter estimates of bus route length in an 800m buffer. The presence of bus shelter has a significant impact on ridership and as expected, the impact is positive on ridership.

Transportation Infrastructure Characteristics

Several transportation infrastructure variables affect boarding and alighting including sidewalk length in a 400m buffer, secondary highway road length in an 800m buffer and local road length in an 800m buffer. All the three variables have a positive impact on ridership. The results are intuitive. In the presence of sidewalk, transit riders have easy access to bus stops and are more likely to use transit in their presence. The presence of secondary highway and local roads is more conducive to transit ridership as opposed to major highways. Also, these roads connect transit riders to the residences (local roads) and potential destinations (secondary highways connecting to jobs and activities).

Built Environment Attributes

The built environment around a bus stop has a significant influence on bus ridership at the stop level. The presence of office area and the institutional area in 800m buffer within a stop significantly increase the bus ridership in Orlando. The proportion of residential area has positive effect on boarding ridership of 800m buffer, but no impact on alighting ridership. On the other hand, the presence of recreation area within 800m buffer of a stop reduces bus ridership. The distance from the CBD from a bus stop negatively impacts the bus ridership. This is expected because ridership is likely to fall as we move away from CBD. As SunRail was introduced during the study period, we also considered the impact of the commuter rail system on bus ridership. In our analysis, we tested several functional forms. Only the inverse distance from SunRail station to the bus stop – referred to as distance decay function - offered significant parameter for ridership. Specifically, the distance decay function with the interaction of SunRail operation period (from March, 2014) is found to have significant impact in the final specified model. The interaction term indicating that the SunRail influence has a negative coefficient implies that the bus ridership is

likely to be lower for the stops closer to the SunRail stations. The result is perhaps indicative of the competition between the two transit systems.

Demographic and Socioeconomic Characteristics

The demographic and socioeconomic variables based on census tract of the bus stop significantly affect bus ridership in Orlando. Specifically, an increased share of rented household in Orlando is likely to increase bus ridership. The automobile ownership variable defined as household with no vehicle has a positive impact on ridership. The result is expected as these individuals are likely to be captive riders and mostly rely on public transportation.

Alternative Specific Effects for Ridership Component

In the ridership components, the alternative specific impacts were examined for a subset of the alternatives (denoted as ρ_j and τ_j). In both components, constant for first two categories offered statistically significant results improving the data fit. These parameters do not have any specific interpretations except to serve as intercepts.

Variance Parameters for Ridership Component

As described in the methodology section, the proposed model framework relaxes the assumption that the error terms ε_{qt} , ξ_{qt} are standard logistically distributed. Hence, the standard deviation of the error terms estimated are reported. While we tested for several variables in the parameterization only the constant variable was statistically significant.

Joint Common Unobserved Factors (Endogeneity)

The last row panel of results in Table 3 provide the estimates for joint common unobserved factors (η'_q) affecting the three dependent variables. The endogeneity impact is captured through two variables – a constant and route length in 800 m buffer. The correlation among the components could be either positive or negative. In our analysis, we found the negative sign to offer better fit for common correlation. Overall, the results clearly support our hypothesis that common unobserved factors influence the three components. The results indicate that as the bus route length increases the influence of common factors affecting the three dependent variables increases.

MODEL VALIDATION

The model developed was validated using a hold-out sample. For this purpose, we generated various measures for the hold-out sample with 644 stops (11 records per stop). We calculated predictive log-likelihood, Bayesian information criterion (BIC), Akaike information criterion (AIC) and Corrected Akaike information criterion (AICc) measures to compare the performance of the estimated independent and joint models. The predictive log-likelihood value for the joint model and independent model are -243,072.04 and -271,401.66, respectively; the corresponding BIC values are 486,640.55 and 543,282.06, respectively. The AIC (AICc) values for the joint and independent models are 486,256.08 (486,256.54) and 542,911.32 (542,911.75) respectively. All of these measures clearly highlight that the improvement in the joint model is not a manifestation of over fitting, and further supports our hypothesis that headway is endogenous to ridership components in the current study context.

POLICY ANALYSIS

To illustrate the influence of headway endogeneity, we investigate change in boarding and alighting ridership categories by changing various independent variables for independent and joint models. The variables considered in our exercise include the number of zero vehicle households, length of sidewalk in a 400m buffer, number of rented households, presence of shelter at the bus stop, change to the headway category, institutional area and residential area in an 800m buffer. For our analysis, we consider a 25% increase in all variables except residential area. For residential area, we consider 10% increase in an 800m buffer variable. For headway variable, we randomly changed headway for 25% of the stops from category 2 to category 3 (i.e. increased headway for some randomly selected stops). Based on these new variable values, an elasticity analysis is performed (see Eluru and Bhat 2007 for a discussion on the methodology for computing elasticities). The results for the elasticity effects across different ridership categories are presented in Table 4 for boardings and alightings for both independent and joint models. The numbers in Table 4 may be interpreted as the percentage change in each ridership categories due to the change in exogenous variable. Several observations can be made from Table 4. First, the elasticity estimates obtained for all variables are significantly different for the independent and the joint model. The elasticity comparison highlights how ignoring the presence of headway endogeneity could result in incorrect estimates and elasticities for headway and other independent variables. Second, sidewalk length, number of rented households and presence of shelter have positive impact on ridership. The changes in these variables is likely to increase the ridership significantly. Third, as expected, change of headway category for category 2 to category 3 has a negative impact on ridership. Finally, the policy analysis illustrates how any potential change to independent variables can be analyzed using the proposed model framework.

The proposed model structure can also be utilized to predict boardings and alightings after making changes to the independent variables. To illustrate this, we perform a forecasting analysis by considering the change in residential area (10% increase), headway (25% increase in headway category 3 to 2), length of sidewalk (25% increase), bike length (25% increase), and zero vehicles household (25% increase). We consider these variables for the forecasting exercise to reflect transit friendly development and investments. In generating the forecasted ridership categories based on the change in above variables, we identify the ridership categories based on probabilistic assignment by using predicted probabilities computed from the joint model. The probabilities are appropriately aggregated across ridership categories to identify the corresponding bin specific frequencies. For illustration purposes, we plot the ridership categories identified for the observed and forecasted scenarios for a sample of 200 randomly selected stops. These plots are presented in Figure 1 (for alighting) and Figure 2 (for boarding). From these figures, it is evident that transit friendly investments have substantial impact on increasing transit ridership providing credence to the rationale that to increase the ridership, service improvements related to public transit (improvement of headway, sidewalk and addition of shelters) should be considered. The development of such forecasting scenarios would assist policy makers and transit agencies in identifying ideal locations for transit friendly infrastructure installation.

Table 4 Policy Analysis

Bin	Zero vehicle HH 25% Increase		Sidewalk 25% increase		Rented HH 25% increase		Presence of Shelter 25% increase		25% add to headway category 2 from headway category 3		Institutional area 25% increase		Residential area 10% increase	
	IM	JM	IM	JM	IM	JM	IM	JM	IM	JM	IM	JM	IM	JM
Boarding Component														
1	-0.49%	-0.57%	-8.44%	-6.19%	-5.37%	-5.66%	-6.78%	-7.74%	4.72%	6.00%	0.57%	-0.14%	-2.07%	-1.67%
2	0.25%	0.22%	3.21%	1.96%	2.76%	2.24%	0.12%	-1.20%	-6.98%	-8.38%	-0.45%	0.09%	2.94%	2.17%
3	0.64%	0.61%	10.72%	6.51%	6.94%	6.06%	8.01%	6.44%	-6.06%	-7.21%	-0.80%	0.16%	3.91%	2.86%
4	0.92%	0.85%	16.60%	9.45%	9.97%	8.40%	15.41%	12.60%	-4.98%	-6.04%	-1.00%	0.20%	3.80%	2.84%
5	1.07%	0.95%	19.80%	10.67%	11.58%	9.35%	19.69%	15.57%	-4.43%	-5.50%	-1.06%	0.22%	3.33%	2.67%
6	1.22%	1.05%	22.82%	11.81%	13.09%	10.22%	23.57%	18.59%	-4.02%	-5.01%	-1.10%	0.23%	2.58%	2.42%
7	1.35%	1.14%	25.55%	12.90%	14.45%	11.07%	26.73%	21.65%	-3.74%	-4.58%	-1.11%	0.24%	1.53%	2.09%
8	1.45%	1.23%	27.75%	13.97%	15.51%	11.92%	28.94%	24.67%	-3.60%	-4.22%	-1.10%	0.25%	0.04%	1.69%
9	1.50%	1.33%	29.11%	15.04%	16.11%	12.80%	30.10%	27.53%	-3.55%	-3.93%	-1.07%	0.26%	-2.12%	1.22%
10	1.51%	1.43%	29.40%	16.09%	16.19%	13.70%	30.17%	30.09%	-3.59%	-3.72%	-1.01%	0.27%	-5.28%	0.65%
11	1.46%	1.52%	28.55%	17.11%	15.79%	14.57%	29.23%	32.27%	-3.73%	-3.57%	-0.93%	0.27%	-9.68%	-0.04%
12	1.34%	1.63%	25.80%	18.35%	14.92%	15.65%	26.78%	34.54%	-4.10%	-3.48%	-0.74%	0.27%	-17.71%	-1.33%
13	0.75%	1.59%	12.07%	17.37%	9.82%	15.68%	15.42%	32.71%	-3.38%	-3.33%	-0.06%	0.21%	-56.61%	-13.87%
Alighting Component														
1	-0.72%	-0.69%	-9.31%	-8.58%	-4.96%	-5.76%	-2.95%	-4.98%	5.04%	6.59%	-0.03%	-0.66%	-	-
2	0.21%	0.14%	1.69%	0.77%	1.87%	1.48%	0.00%	-0.88%	-6.86%	-8.50%	0.02%	0.26%	-	-
3	0.82%	0.64%	10.00%	7.33%	5.74%	5.45%	2.93%	3.40%	-5.87%	-7.24%	0.04%	0.67%	-	-
4	1.30%	0.97%	17.25%	12.06%	8.67%	7.97%	5.89%	7.00%	-4.72%	-5.94%	0.05%	0.91%	-	-
5	1.58%	1.11%	21.72%	14.20%	10.31%	9.02%	7.87%	8.91%	-4.14%	-5.31%	0.05%	1.00%	-	-
6	1.84%	1.25%	26.22%	16.26%	11.82%	10.02%	9.78%	10.95%	-3.70%	-4.75%	0.06%	1.07%	-	-
7	2.07%	1.39%	30.55%	18.36%	13.11%	11.02%	11.43%	13.12%	-3.42%	-4.27%	0.06%	1.15%	-	-

8	2.25%	1.53%	34.34%	20.56%	14.10%	12.05%	12.68%	15.35%	-3.25%	-3.88%	0.06%	1.22%	-	-
9	2.37%	1.67%	37.22%	22.89%	14.76%	13.09%	13.50%	17.52%	-3.19%	-3.57%	0.06%	1.28%	-	-
10	2.44%	1.81%	38.96%	25.30%	15.05%	14.12%	13.86%	19.52%	-3.17%	-3.33%	0.06%	1.35%	-	-
11	2.43%	1.95%	39.44%	27.68%	14.97%	15.08%	13.77%	21.26%	-3.14%	-3.16%	0.06%	1.40%	-	-
12	2.30%	2.10%	37.70%	30.65%	14.16%	16.18%	12.77%	23.13%	-2.98%	-3.01%	0.06%	1.46%	-	-
13	1.76%	2.20%	24.94%	32.77%	11.18%	16.80%	10.11%	24.00%	-1.97%	-2.67%	0.03%	1.40%	-	-
<i>Note: IM = Independent Model; JM = Joint Model</i>														

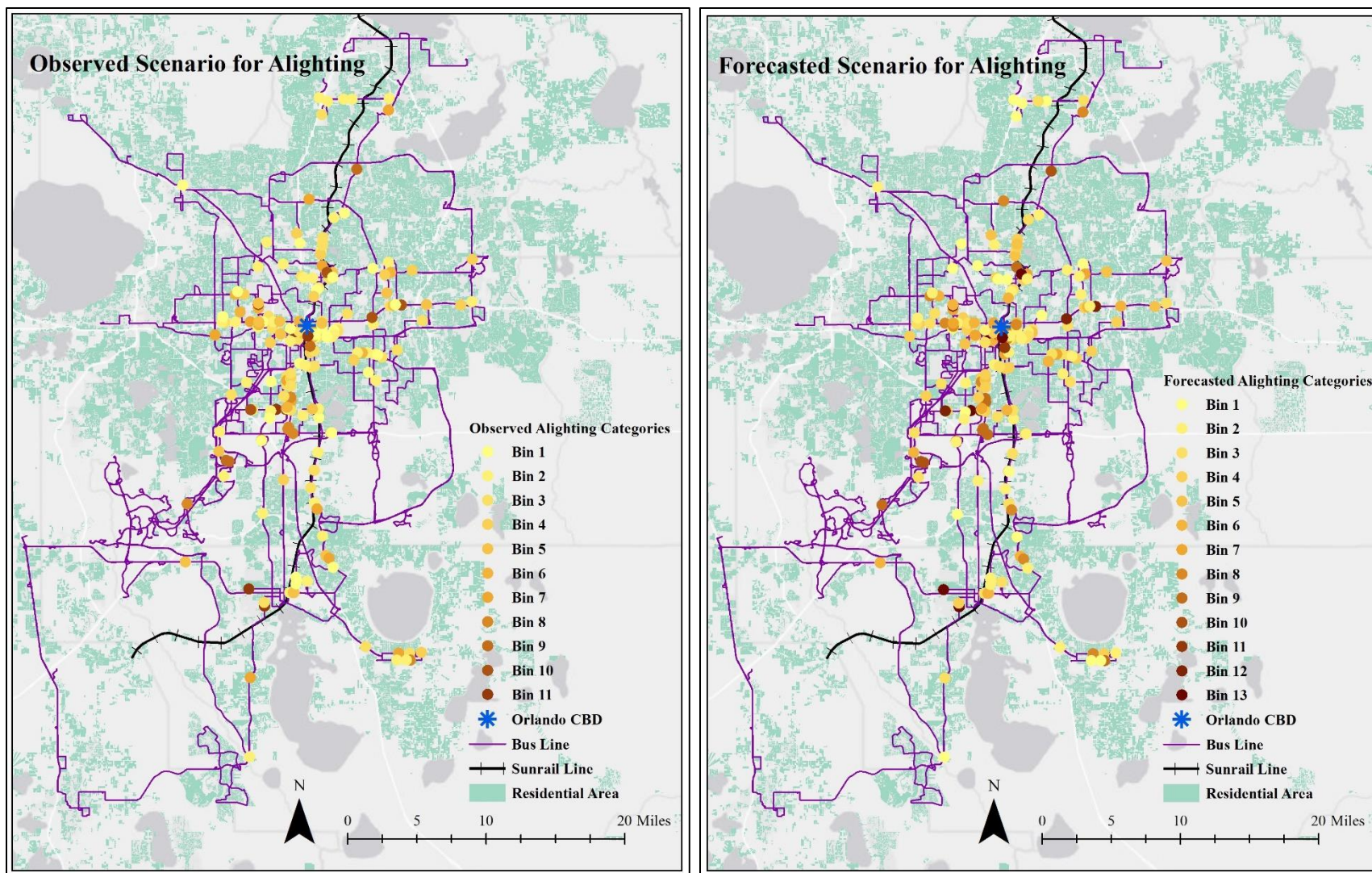


Figure 1 Policy Analysis (Observed and Forecasted Alighting Ridership Categories)

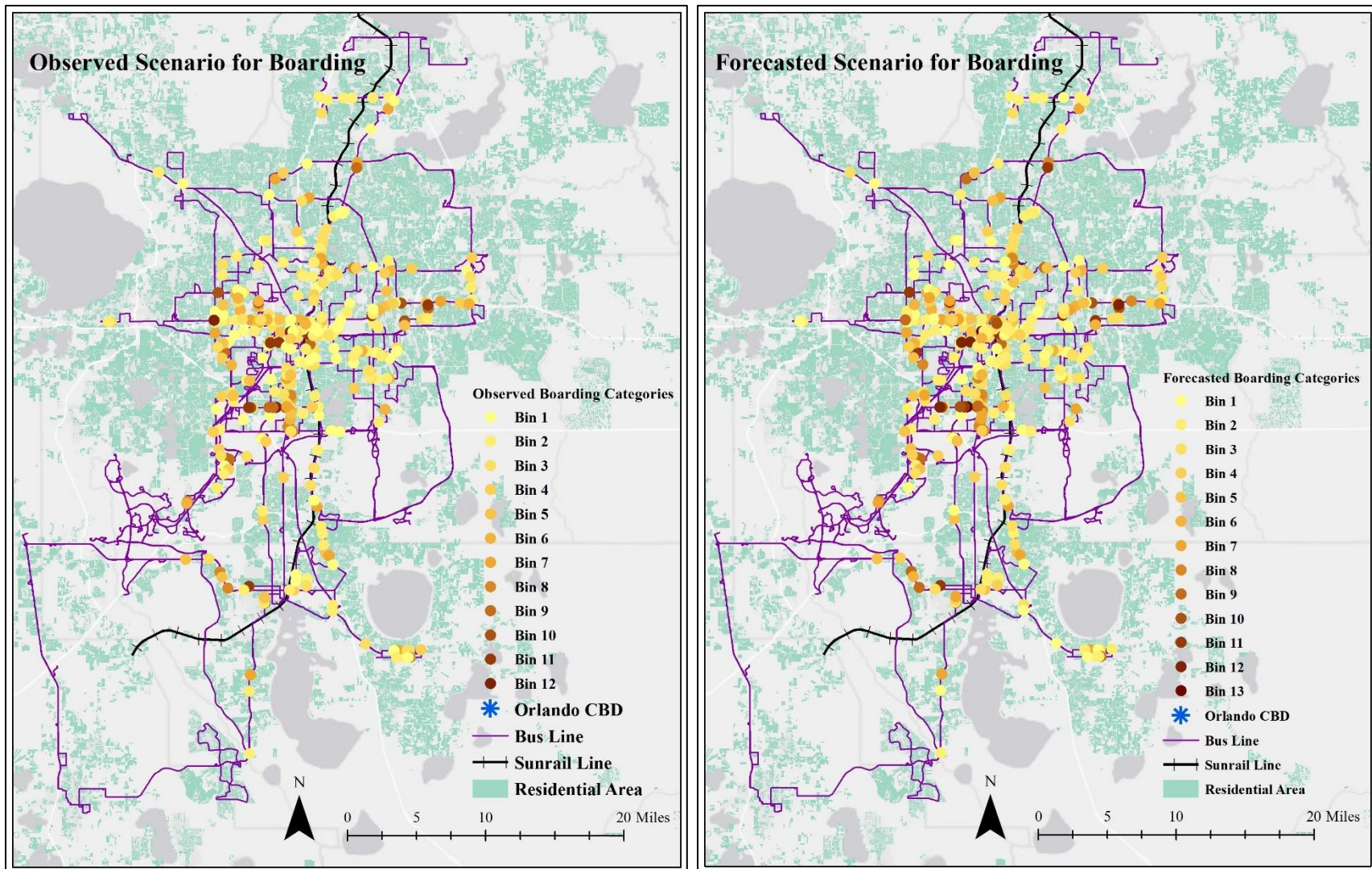


Figure 2 Policy Analysis (Observed and Forecasted Boarding Ridership Categories)

CONCLUSIONS

Policy makers and urban agencies, across different parts of North America, are considering investments in various public transportation alternatives including bus, light rail, commuter rail, and metro. A critical component to evaluating the success of these investments is the development of appropriate statistical tools to examine the impact. Our proposed research contributes to public transit literature by developing econometric models that consider the potential endogeneity of stop level headway in modeling ridership. Most of the earlier studies in public transit literature ignore that the stop level headway was determined (by choice) in response to the expected ridership. In traditional ridership studies, this potential endogeneity is often neglected, and headway is considered as an independent variable. The approach violates the requirement that the unobserved factors that affect the dependent variable do not affect the independent variable. If this is the case, the estimated impact of headway on ridership would be biased (potentially over-estimated). More importantly, the estimated impact of all other variables (such as land use factors, bus infrastructure) will also be biased (possible under-estimated).

In this study, we addressed these challenges by proposing a simultaneous equation system that considered headway and ridership in a joint framework accounting for the influence of common unobserved factors that affect headway and ridership. The proposed model was developed employing ridership data from the Lynx bus transit system of the Greater Orlando region. The ridership data included stop level average weekday boarding and alighting information for 11 quadrimesters from May-2013 to December-2016. The presence of multiple data points for each stop allowed us to develop panel models for headway, boarding and alighting. The headway variable was modeled using a panel ordered logit model while the ridership variables were modeled using panel grouped ordered logit models. In addition to unobserved effects in the form panel random effects, several exogenous variables including stop level attributes, transportation infrastructure variables, transit infrastructure variables, land use and built environment attributes and demographic and socioeconomic variables in the vicinity of the bus stop were considered in the model estimation.

The proposed model was estimated and compared to the model that ignore the potential endogeneity of headway. The comparison results clearly highlighted the improved model fit of the proposed framework. The model estimation results identified that headway, number of the bus stops in the 800m buffer, presence of shelter at the bus stop, sidewalk length in a 400m buffer, bus stop distance from the central business district (CBD), distance between Sunrail station and bus stop, and automobile ownership are likely to impact bus ridership in Orlando. The model system developed was also validated with a hold-out sample. Finally, a policy analysis exercise was conducted to illustrate how the model framework can be used by policy makers. The model developed can provide policy makers the necessary framework to evaluate changes to ridership based on a series of changes under consideration.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Mr. William Gant from Lynx and Mr. Jeremy Dilmore from Florida Department of Transportation (FDOT) for helping with data acquisition. The authors would also like to thank FDOT for funding this study. All opinions are only those of the authors.

REFERENCES

- Abay, K.A., Paleti, R., Bhat, C.R., 2013. The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. *Transportation Research Part B: Methodological* 50, 74-89.
- Aditjandra, P.T., Cao, X., Mulley, C., 2012. Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a british metropolitan data. *Transportation Research Part A: Policy and Practice* 46 (1), 22-32.
- Babalik-Sutcliffe, E., 2002. Urban rail systems: Analysis of the factors behind success. *Transport Reviews* 22 (4), 415-447.
- Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., White, P., 2004. The demand for public transport: A practical guide.
- Beck, M.J., Rose, J.M. and Greaves, S.P., 2017. I can't believe your attitude: a joint estimation of best worst attitudes and electric vehicle choice. *Transportation*, 44(4), pp.753-772.
- Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences. *Transportation Research Part B: Methodological* 37 (9), 837-855.
- Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B: Methodological* 43 (7), 749-765.
- Bhat, C.R., Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological* 41 (5), 506-526.
- Bliss, L., 2017. What's behind declining transit ridership nationwide? Citylab, February 24, 2017.
- Boarnet, M.G., Giuliano, G., Hou, Y., Shin, E.J., 2017. First/last mile transit access as an equity planning issue. *Transportation Research Part A: Policy and Practice* 103, 296-310.
- Cao, X., Mokhtarian, P.L., Handy, S.L., 2009. Examining the impacts of residential self-selection on travel behaviour: A focus on empirical findings. *Transport Reviews* 29 (3), 359-395.
- Chakour, V., Eluru, N., 2014. Analyzing commuter train user behavior: A decision framework for access mode and station choice. *Transportation* 41 (1), 211-228.
- Chorus, C. G., & Kroesen, M. (2014). On the (im-) possibility of deriving transport policy implications from hybrid choice models. *Transport Policy*, 36, 217-222.
- Creutzig, F., 2014. How fuel prices determine public transport infrastructure, modal shares and urban form. *Urban Climate* 10, 63-76.
- Debrezion, G., Pels, E., Rietveld, P., 2007. The impact of railway stations on residential and commercial property value: A meta-analysis. *The Journal of Real Estate Finance and Economics* 35 (2), 161-180.
- Debrezion, G., Pels, E., Rietveld, P., 2009. Modelling the joint access mode and railway station choice. *Transportation Research Part E: logistics and transportation review* 45 (1), 270-283.
- Deka, D., 2002. Transit availability and automobile ownership: Some policy implications. *Journal of Planning Education and Research* 21 (3), 285-300.
- Ding, C., Wang, D., Liu, C., Zhang, Y., Yang, J., 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice* 100, 65-80.
- Eluru, N., Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis & Prevention* 39 (5), 1037-1049.

- Ettema, D., Nieuwenhuis, R., 2017. Residential self-selection and travel behaviour: What are the effects of attitudes, reasons for location choice and the built environment? *Journal of Transport Geography* 59, 146-155.
- Evans, I., 2004. Transit scheduling and frequency-traveler response to transportation system changes.
- Faghih-Imani, A., Eluru, N., 2016. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of new york citibike system. *Journal of Transport Geography* 54, 218-227.
- Fan, K.-S., Miller, E.J., Badoe, D., 1993. Modeling rail access mode and station choice. *Transportation Research Record* (1413).
- Fernández-Antolín, A., Guevara, C.A., De Lapparent, M. and Bierlaire, M., 2016. Correcting for endogeneity due to omitted attitudes: Empirical assessment of a modified MIS method using RP mode choice data. *Journal of choice modelling*, 20, pp.1-15.
- Fitzroy, F., Smith, I., 1998. Public transport demand in freiburg: Why did patronage double in a decade? *Transport policy* 5 (3), 163-173.
- Fitzroy, F., Smith, I., 1999. Season tickets and the demand for public transport. *Kyklos* 52 (2), 219-238.
- Garrett, M., Taylor, B., 1999. Reconsidering social equity in public transit. *Berkeley Planning Journal* 13 (1).
- Gomez-Ibanez, J.A., 1996. Big-city transit rider snip, deficits, and politics: Avoiding reality in boston. *Journal of the American Planning Association* 62 (1), 30-50.
- Grengs, J., 2010. Job accessibility and the modal mismatch in detroit. *Journal of Transport Geography* 18 (1), 42-54.
- Guevara, C.A. and Ben-Akiva, M.E., 2012. Change of scale and forecasting with the control-function method in logit models. *Transportation Science*, 46(3), pp.425-437.
- Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part A: Policy and Practice*, 82, pp.240-254.
- Guevara, C.A. and Polanco, D., 2016. Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. *Transportmetrica A: Transport Science*, 12(5), pp.458-478.
- Handy, S., 1996. Methodologies for exploring the link between urban form and travel behavior. *Transportation Research Part D: Transport and Environment* 1 (2), 151-165.
- Handy, S., Cao, X., Mokhtarian, P., 2005. Correlation or causality between the built environment and travel behavior? Evidence from northern california. *Transportation Research Part D: Transport and Environment* 10 (6), 427-444.
- Heckman, J.J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica* 73 (3), 669-738.
- Holmgren, J., 2007. Meta-analysis of public transport demand. *Transportation Research Part A: Policy and Practice* 41 (10), 1021-1035.
- Jaffe, E., 2014. A basic shelter can make the wait for the bus feel shorter.
- Johnson, A., 2003. Bus transit and land use: Illuminating the interaction. *Journal of Public Transportation* 6 (4), 2.
- Kain, J.F., Liu, Z., 1999. Secrets of success: Assessing the large increases in transit ridership achieved by houston and san diego transit providers. *Transportation Research Part A: Policy and Practice* 33 (7), 601-624.

- Kamargianni, M., Ben-Akiva, M. and Polydoropoulou, A., 2014. Incorporating social interaction into hybrid choice models. *Transportation*, 41(6), pp.1263-1285.
- Kawabata, M., Shen, Q., 2006. Job accessibility as an indicator of auto-oriented urban structure: A comparison of boston and los angeles with tokyo. *Environment and Planning B: Planning and Design* 33 (1), 115-130.
- Kawabata, M., Shen, Q., 2007. Commuting inequality between cars and public transit: The case of the san francisco bay area, 1990-2000. *Urban Studies* 44 (9), 1759-1780.
- Kim, J., Rasouli, S. and Timmermans, H., 2014. Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: Application to intended purchase of electric cars. *Transportation research part A: policy and practice*, 69, pp.71-85.
- Kim, J., Rasouli, S. and Timmermans, H.J., 2017. The effects of activity-travel context and individual attitudes on car-sharing decisions under travel time uncertainty: A hybrid choice modeling approach. *Transportation Research Part D: Transport and Environment*, 56, pp.189-202.
- Lewyn, M., 2018. Why is transit ridership declining?
- Ma, X., Chen, X., Li, X., Ding, C., Wang, Y., 2018. Sustainable station-level planning: An integrated transport and land use design model for transit-oriented development. *Journal of Cleaner Production* 170, 1052-1063.
- Mackett, R., Sutcliffe, E.B., 2003. New urban rail systems: A policy-based technique to make them more successful. *Journal of Transport Geography* 11 (2), 151-164.
- Maddala, G.S., 1983. Limited-dependent and qualitative variables in econometrics Cambridge university press.
- Mariel, P., Hoyos, D., Artabe, A. and Guevara, C.A., 2018. A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation. *Science of the Total Environment*, 633, pp.967-980.
- Mccollom, B.E., Pratt, R.H., 2004a. Traveler response to transportation system changes handbook, third edition: Chapter 12, transit pricing and fares The National Academies Press, Washington, DC.
- Mccollom, B.E., Pratt, R.H., 2004b. Traveler response to transportation system changes. Chapter 12-transit pricing and fares.
- Mokhtarian, P.L., Cao, X., 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological* 42 (3), 204-228.
- Nowak, W.P., Savage, I., 2013. The cross elasticity between gasoline prices and transit use: Evidence from chicago. *Transport Policy* 29, 38-45.
- Paulssen, M., Temme, D., Vij, A. and Walker, J.L., 2014. Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. *Transportation*, 41(4), pp.873-888.
- Pinjari, A.R., Bhat, C.R., Hensher, D.A., 2009. Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B: Methodological* 43 (7), 729-748.
- Rahman, M., Yasmin, S., & Eluru, N. (2019). Evaluating the impact of a newly added commuter rail system on bus ridership: a grouped ordered logit model approach. *Transportmetrica A: Transport Science*, 1-21.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2), 135-146.

- Schmitt, A., 2017. Transit ridership falling everywhere — but not in cities with redesigned bus networks.
- Seskin, S., Cervero, R., Zupan, J., Howard, J., 1996. Transit and urban form. Federal Transit Administration, Washington, DC.
- Shen, Q., 2001. A spatial analysis of job openings and access in a u.S. Metropolitan area. *Journal of the American Planning Association* 67 (1), 53-68.
- Siddiqui, F., 2018. Falling transit ridership poses an ‘emergency’ for cities, experts fear.
- Tirachini, A., Hensher, D.A. and Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation research part A: policy and practice*, 53, pp.36-52.
- Tomer, A., 2016. America’s commuting choices: 5 major takeaways from 2016 census data.
- Tp, 2016. Transport politic projects under consideration summary.
- Van Acker, V., Van Wee, B., Witlox, F., 2010. When transport geography meets social psychology: Toward a conceptual model of travel behaviour. *Transport Reviews* 30 (2), 219-240.
- Vij, A., Walker, J.L., 2014. Preference endogeneity in discrete choice models. *Transportation Research Part B: Methodological* 64, 90-105.
- Voith, R., 1991. The long-run elasticity of demand for commuter rail transportation. *Journal of Urban Economics* 30 (3), 360-372.
- Voith, R., 1997. Fares, service levels, and demographics: What determines commuter rail ridership in the long run? *Journal of Urban Economics* 41 (2), 176-197.
- Walker, J.L., Ehlers, E., Banerjee, I., Dugundji, E.R., 2011. Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A: Policy and Practice* 45 (4), 362-374.
- Wardman, M., Whelan, G., 1999. Using geographical information systems to improve rail demand models. Final report to Engineering and Physical science Research council.
- Wardman, M. and Whelan, G., 2011. Twenty years of rail crowding valuation studies: evidence and lessons from British experience. *Transport reviews*, 31(3), pp.379-398.
- Wenglenski, S., Orfeuill, J.-P., 2004. Differences in accessibility to the job market according to social status and place of residence in the paris area. *Built Environment* 30 (2), 116-126.
- Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis & Prevention* 59, 506-521.
- Yasmin, S., Eluru, N., Pinjari, A.R., 2015. Analyzing the continuum of fatal crashes: A generalized ordered approach. *Analytic Methods in Accident Research* 7, 1-15.

APPENDIX A Lynx Ridership Independent Model Results –

Variable Name	Headway		Alighting		Boarding	
	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
Independent Variables (v, α, β)						
Constant	-	-	-12.06	-22.96	-18.65	-27.88
Threshold Value 1	-2.20	-66.78	-	-	-	-
Threshold Value 2	0.40	12.49	-	-	-	-
<i>Stop Level Attributes</i>						
Headway (Base: Category 1)						
<i>Dummy for headway category 2</i>	-	-	-22.77	-70.01	-25.99	-71.63
<i>Dummy for headway category 3</i>	-	-	-39.64	-107.02	-42.69	-103.36
No of Bus stop in an 800 m buffer	-	-	-2.41	-27.99	-2.77	-28.14
Bus route Length in an 800 m buffer	-0.44	-66.26	-0.70	-6.99	-0.80	-6.59
Presence of shelter in bus stop	-	-	7.77	44.99	19.71	97.05
<i>Transportation Infrastructure</i>						
Sidewalk length in an 400 m buffer	-	-	1.97	17.55	1.98	15.79
Bike road length in an 800 m buffer	-0.17	-38.02	-	-	-	-
Secondary road length in an 800 m buffer	-0.32	-42.02	5.55	40.65	4.72	31.61
Local road length in an 800 m buffer	0.26	27.24	1.87	14.34	1.32	8.00
Railroad length in an 800 m buffer	-0.27	-37.84	-	-	-	-
<i>Built Environment and Land-use Attributes</i>						
Land use area type in an 800m buffer						
<i>Institutional area</i>	-1.68	-21.10	0.75	0.58	-16.81	-11.14
<i>Residential area</i>	1.36	39.35	-	-	11.91	20.06
<i>Office area</i>	-1.89	-37.55	30.93	50.71	25.94	28.63
<i>Recreational area</i>	-1.37	-11.11	-46.26	-25.16	-39.75	-19.09
<i>Industrial Area</i>	2.63	42.87	-	-	-	-
Distance from Central business district (CBD)	0.59	78.50	-1.09	-9.37	-1.98	-15.20
<i>Demographic and Socioeconomic Variables</i>						
Zero vehicle in HH	-1.37	-14.36	53.86	32.60	41.89	22.30
High income population	-0.67	-14.67	-	-	-	-
Household rent	-	-	17.90	43.25	22.24	47.32
<i>SunRail Effects</i>						
Distance Decay Function for SunRail*SunRail operation period	-	-	-5.06	-27.50	-5.35	-24.74
Alternative Specific Effects (ρ_j and τ_j)						
Constant – Alternative 1 (0-5 ridership)	-	-	22.63	90.62	25.89	86.68

Constant – Alternative 2 (>5-10 ridership)	-	-	10.61	58.69	12.29	57.54
Variance (λ_B^2 and λ_A^2)						
Constant	-	-	2.86	571.65	2.92	516.67